Big Data Ethics: A Life Cycle Perspective

Simon Vydra, Andrei Poama, Sarah Giest, Alex Ingrams & Bram Klievink*

Abstract

The adoption of big data analysis in the legal domain is a recent but growing trend that highlights ethical concerns not just with big data analysis, as such, but also with its deployment in the legal domain. This article systematically analyses five big data use cases from the legal domain utilising a pluralistic and pragmatic mode of ethical reasoning. In each case we analyse what happens with data from its creation to its eventual archival or deletion, for which we utilise the concept of 'data life cycle'. Despite the exploratory nature of this article and some limitations of our approach, the systematic summary we deliver depicts the five cases in detail, reinforces the idea that ethically significant issues exist across the entire big data life cycle, and facilitates understanding of how various ethical considerations interact with one another throughout the big data life cycle. Furthermore, owing to its pragmatic and pluralist nature, the approach is potentially useful for practitioners aiming to interrogate big data use cases.

Keywords: big data, big data analysis, data life cycle, ethics, AI

1 Introduction

The transformative potential of big data has attracted considerable academic attention in the last two decades, focusing not only on developments in the private sector, but also on big data's impact on various aspects of governance and public policy. As this special issue demonstrates, big data analytics are also becoming more widely used in the legal domain – a development that raises new ethical questions and concerns. In the legal context, key concerns have to do with judicial and legal principles and can be a lot more controversial than in contexts where model performance is the key (or perhaps the only) relevant metric.

This article explores how the use of big data analytics in the legal domain raises moral questions, looking closely at five illustrative cases and doing so in a structured systematic way. Our approach here is informed by the fact that many ethically significant decisions arise either before analysis, when collecting, storing, and aggregating data, or after analysis, when results are communicated, decisions are reached and lessons are learned. Thus, examining the morality of big data use in a systematic way requires that we include and consider the moral dimensions of all stages of the process, for which we utilise the 'data life cycle' concept. To interrogate the morality of big data use along the various stages of a data life cycle we adopt a 'lawyerly' mode of ethical reasoning that is ethically pluralistic and pragmatic in the sense of arriving at a convincing ethical argument for or against a given practice. The contribution of the article is thus twofold: First, it summarises five big data uses cases from the legal domain in a structured way, pointing to details that might not be obvious otherwise. Secondly, it proposes an approach of morally interrogating big data use cases that combines the logic of following the data's life cycle with a 'lawyerly' perspective, making it potentially valuable to those aiming to interrogate (and change) big data systems in practice.

This article is structured as follows. In Section 1, we articulate a framework for examining big data ethics as applied to the legal domain. We do this by first defining 'big data' (Section 1.1) and specifying the big data life cycle model on which we rely to articulate a systematic view of big data ethics (Section 1.2). Also, the article advances a working conception of the type of considerations deemed 'ethical' in the context of big data ethics (Section 1.3). Section 2, which makes up the bulk of the article, first introduces the five illustrative cases and then proceeds to briefly describe each case and illustrate its ethical significance for the stages of the big data life cycle. In Section 3 we bring together the insights gained through the different cases to offer a systematic overview of some of the key ethical concerns that might arise along the big data life cycle. Section 4 notes some limitations and concludes the discussion.

1.1 Big Data

The term 'big data' is conceptually fuzzy. The industry-standard definition of 'big data' uses a set of 'Vs': attributes of a data set that all begin with 'V', most commonly volume, variety, velocity and veracity.[1] In academic writing, volume, variety and velocity are used most commonly.[2] Some authors expand on this list by includ-

* Simon Vydra is a Researcher at the Institute for Public Administration, Leiden University, the Netherlands. Andrei Poama is Assistant Professor at the Institute for Public Administration, Leiden University, the Netherlands. Sarah Giest is Assistant Professor at the Institute for Public Administration, Leiden University, the Netherlands. Alex Ingrams is Assistant Professor at the Institute for Public Administration, Leiden University, the Netherlands. Bram Klievink is Professor of Digitization and Public Policy at the Institute for Public Administration, Leiden University, the Netherlands.

1. IBM, '4 Vs', www.ibmbigdatahub.com/tag/587 (2012); J.S. Ward and A. Barker, 'Undefined By Data: A Survey of Big Data Definitions', http://arxiv.org/abs/1309.5821 (2013).
2. O. Ylijoki and J. Porras, 'Perspectives to Definition of Big Data: A Mapping Study and Discussion', 4(1) *Journal of Innovation Management* 69, at 79 (2016).

ing veracity, variability, visualisation and value.[3] Defining big data using a set of 'Vs' allows for simple categorisation but lacks a threshold on these 'Vs' when 'data' become 'big data'. These thresholds are not only somewhat subjective, but also constantly changing as a result of technological advancements.

An alternative approach adopted in this article is to partially avoid these issues by focusing on the overall process and analytics necessary to use this data.[4] Using this approach enables us to focus on the processes linked to utilising data instead of focusing on differences in the data itself. For such a definition this article uses the work of Klievink et al. (2017), who distil a set of five criteria from the available literature:

1. Use and combining of multiple, large datasets, from various sources, both *external and internal* to the organization;
2. Use and combining of *structured* (traditional) and *less structured or unstructured* (nontraditional) data in analysis activities;
3. Use of incoming data streams in *real time* or near real time;
4. Development and application of *advanced analytics and algorithms*, distributed computing and/or advanced technology to handle very large and complex computing tasks;
5. *Innovative use* of existing datasets and/or data sources for new and radically different applications than the data were gathered for or spring from.[5]

This definition remains fuzzy because defining 'advanced analytics' or 'innovative use' remains subjective; however, it captures an important aspect of big data crucial for this article: the fact that big data is often 're-purposed' data that was not originally intended for the analysis it is being used for. It also aligns with the data life cycle perspective adopted by this article.

1.2 Big Data Life Cycle

The ambition of the data life cycle concept is to 'present a structure for organising the tasks and activities related to the management of data within a project or an organization'.[6] The concept is operationalised into various data life cycle models that cover the entire life of (big) data from generation to archiving/deletion and views the entire process as feeding into the next iteration of the same process, making it a cycle. What makes such a concept crucial for this article is that it 'provide[s] a structure for considering the many operations that will need to be performed on a data record throughout its life'.[7] Ethically significant decisions are not limited to the stage of generating insight and using it. Operations performed on/with data that precede and follow the decision-making step are no less ethically significant, and to conduct a thorough review of the ethical aspects of big data use every step of the cycle should be considered.[8]

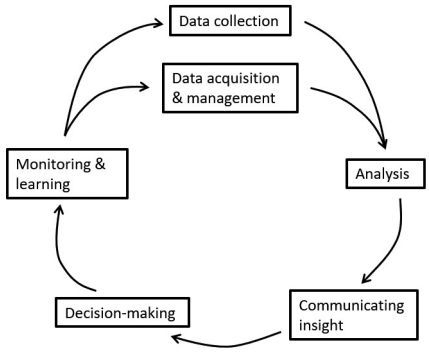
One particular challenge of a life cycle approach is that there is no unified (big) data life cycle model as data life cycles are very different per domain, field and even organisation.[9] Although there have been attempts to develop a scenario-agnostic data life cycle model with a broad application,[10] they do not capture big data use in the legal domain well enough. Approaches that adapt the data life cycle to big data make useful changes but are based on a 'Vs' definition of big data and are not specific to the legal domain.[11] This forces us to select a particular life cycle model, and in doing so we face an important trade-off between generality and complexity: the more complex a data life cycle model is, the better it describes an individual case but lacks generality for describing other big data use cases. This is important because in the legal domain different uses of data can correspond to different life cycles: data can just be archived for record-keeping, can be used for a one-off lawmaking decision or can be continuously processed in a decision-support system. When interrogating a singular big data use case it is, of course, reasonable to follow a data life cycle model that fits that case as much as possible – we recognise that generality is more a feature of social scientific inquiry than a feature of the legal process. In this article we aim to maintain a level of generality to illustrate the merit of our selected approach across multiple cases and to be able to articulate some more general conclusions about ethical concerns with big data use cases in the legal domain.

The data life cycle model we adopt is described in Figure 1, and it aims to be simple enough to generalise across many different use cases in the legal domain but also specific enough to capture meaningful and distinct 'stages'. In this data life cycle model we include six distinct stages: the collection of data, which can involve both actively seeking and storing information and more passive collection of information of no obvious analytical value at the time of collection. The acquisition of that data, which entails purchasing or otherwise obtaining data that is already collected by another actor to either

3. *Ibid.*
4. Approach that is arguably similar to that of G.H. Kim, S. Trimi, and J.H. Chung, 'Big-Data Applications in the Government Sector', 57(3) *Communications of the ACM* 78 (2014).
5. B. Klievink, B.J. Romijn, S. Cunningham, and H. de Bruijn, 'Big Data in the Public Sector: Uncertainties and Readiness', 19(2) *Information Systems Frontiers* 267 ,at 269 (2017).
6. L. Pouchard, 'Revisiting the Data Lifecycle with Big Data Curation', 10(2) *International Journal of Digital Curation* 176, at 180 (2016).
7. A. Ball, 'Review of Data Management Lifecycle Models', https://researchportal.bath.ac.uk/en/publications/review-of-data-management-lifecycle-models (2012), at 4.
8. J.M. Wing, 'The Data Life Cycle', 1(1) *Harvard Data Science Review* 1, at 4 (2019).
9. Ball, above n. 7; M. El Arass, I. Tikito, and N. Souissi, 'Data Lifecycles Analysis: Towards Intelligent Cycle', *Intelligent Systems and Computer Vision ISCV 2017* (2017); Pouchard, above n. 6; A. Sinaeepourfard, J. Garcia, X. Masip-Bruin, and E. Marín-Torder, 'Towards a Comprehensive Data LifeCycle Model for Big Data Environments', in *Proceedings of the 3rd IEEE/ACM International Conference on Big Data Computing, Applications and Technologies* (2016).
10. Sinaeepourfard et al., above n. 9.
11. Pouchard, above n. 6; Sinaeepourfard et al., above n. 9.

further utilise the data on its own or to join it with already collected/acquired data. The analysis of that data, which includes cleaning and processing of data to extract conclusions that are valuable for various types of decision-making. The communication of these conclusions, which involves the selection of 'useful' conclusions and techniques (such as data visualisation) aimed at reducing complexity and information overload associated with communicating these conclusions to (technically non-expert and time constrained) decision-makers. And, finally, monitoring and learning, which includes interrogating (internally or externally) the outcomes and taking corrective or optimising steps for the next iteration of the cycle.[12] Not all of these stages have to be carried out by a single actor (e.g. since big data is often considered to be repurposed, the data is often collected by a different actor) in order to justify the moral relevance of utilising the date.

Figure 1 Big data life cycle[13]



This structure not only allows us to conduct a structured review, but also outlines a potential starting point for practitioners and legal researchers to interrogate real-world big data use cases in terms of ethical concerns. As mentioned previously, for assessing individual cases the big data life cycle model can be much more specific, but even for other studies striving for a degree of generality the individual stages can (and should) be different from those we select for this article. In general, data life cycle models mention very similar stages but name them differently and attach them to or detach them from each other differently,[14] meaning that the version of it we adopt here is not the only justifiable one. This includes how decision-making phases or management activities are included in the model (if at all) – in our case those phases are included and considered important. Such decisions are made in an attempt to balance the generality and complexity of the resulting model, but decisions on this trade-off are inherently subjective and fit for purpose.

1.3 Ethical Significance

In each stage of the data life cycle we aim to describe relevant ethical considerations, requiring a definition of ethical considerations that distinguishes them from any other class of considerations. To do so, we propose and adopt a 'lawyerly' perspective of ethical reasoning. In doing this, we draw on an analogy philosophers sometimes use to distinguish between (currently two) different perspectives that might inform and explain how we reason ethically: the lawmaker's perspective and the judge's perspective.[15] The lawmaker's or politician's perspective focuses on devising laws and policies to increase their constituents' aggregate well-being. This is the practical perspective usually at work in consequentialist ethical theories: thinking like an ideal lawmaker means thinking about how to maximise a particular moral value – typically, utility, but also solidarity, community or care. The judge's perspective, on the other hand, focuses on solving a conflict in a specific case according to a fixed set of rules. This is the practical perspective usually at work in deontological ethical theories: thinking like an ideal judge means thinking in terms of respecting the constraints and prohibitions posited by a specific rule (or set of rules).

Drawing on this role analogy, we advance a third practical perspective: the lawyer's perspective. Unlike the lawmaker's or the judge's perspective, the lawyer's practical perspective focuses on convincingly contesting or defending an action or practice on the basis of a given ethical consideration. The focus is thus not on the structure or substance of the ethical consideration that guides the lawyer's contestation or defence, but rather on the consideration that the 'ethical case' they make is persuasive and sound. Construed from the lawyer's perspective, the point of ethical reasoning is to produce winning ethical arguments – i.e. arguments that have a concrete practical bearing and contribute to making a change in individuals' lives. The lawyerly mode of reasoning can thus be considered ethically pluralistic and opportunistic,[16] in the sense that it remains pragmatically open to a plurality of ethical considerations, and focuses on the values and principles that are most fitting for making a successful argument in a specific case. Although this is merely a cursory description of how the lawyerly mode of ethical reasoning might work, many lawyers might admit that partiality, adversariality and pragmatism are

12. More complexity can be added to this model by including tasks that iteratively happen in every stage, such as the 'describe' and 'assure' steps proposed by Pouchard, above n. 6 to guarantee data quality. Such additional steps support the overall ambition of this article to interrogate big data use at every stage.
13. Authors' own diagram.
14. M. El Arass & N. Souissi, 'Data Lifecycle: From Big Data to SmartData', 2018 *Colloquium in Information Science and Technology* 80, at 80-81 (2018).
15. R. Hare, *Moral Thinking: Its Levels, Method, and Point* (1981); J. Rawls, 'Two Concepts of Rules', 64(1) *The Philosophical Review* 1, at 6 (1955). Here Rawls famously suggests that 'different sorts of arguments are suited to different offices. One way of taking the differences between ethical theories is to regard them as accounts of the reasons expected in different offices'.
16. On moral opportunism, *see* K. Winston, 'Moral Opportunism: A Case Study', 40 *NOMOS: American Society for Political and Legal Philosophy* 154 (1998).

central to how they reason and argue qua lawyers.[17] It is also a mode of ethical reasoning compatible with moral psychology theories that argue that when supporting an ethical position, human reason is more analogous to a lawyer defending a partial position than to an impartial judge applying general rules (Haidt 2012).

Assuming this perspective, we require a definition of 'ethical concerns' that is not committed to a specific ethical theory or a moral outlook and one that is based on guidelines that are more likely to influence big data policies and use cases in the real world (as compared with ethical theories that do not go beyond the confines of academia). Consequently, we distinguish between ethical concerns and non-ethical concerns on the basis of widely shared legal and policy documents. Ethical considerations contained in those documents are considered 'ethical considerations' for the purposes of this article. This allows us, as our lawyerly perspective requires, to avoid stipulating or engaging in normative arguments about what ultimately counts as an ethically significant consideration. Our choice here is pragmatic, in that we defer to those organisations whose recommendations are either binding or widely accepted (in the European context) to determine what an ethically significant consideration is without endorsing any particular ethical viewpoint ourselves. This allows us to avoid the daunting task of finding a common denominator to radically opposed ethical theories and also to rely on a rough-and-ready 'consensus' and a 'practice-informed' account of the considerations that matter ethically.

This definition naturally begs two questions: which documents are considered and how are relevant ethical considerations extracted from them? In terms of the documents considered, we select for documents that are a) either legal prescriptions or government-endorsed recommendations that have practical traction in the sense that big data use case should be reasonably expected to take them into account b) applicable in the European context, and c) concerned either with general research conduct or with advanced analytical methods often used for big data analysis (such as artificial intelligence (AI) or machine learning). The documents meeting these criteria range from EU-level legislation such as the General Data Protection Regulation (GDPR) to more global recommendations endorsed by a majority of European countries (such as recommendations of the UN or the OECD). The sampling of these documents is purposive, providing a good breadth of various types of documents but not assuring that our selection of documents is exhaustive.

In terms of extracting ethical considerations from these documents, the method is conventional content analysis (in the sense that the coding scheme is inferred from the documents themselves) and is focused on explicit declarations of ethical principles within the selected documents. In other words, either the selected documents have a statement of ethical principles that guide the recommendations or the entire document is a list of ethical principles to be followed. Given that these are often listed as individual points or principles, their extraction from the text is very straightforward. Our treatment of the individual ethical considerations extracted from these documents is also rather simple: we list them in a table together with the documents that mention them and whether these documents are about general research conduct, big data use, or both (Annex A). In compiling this table our commitment to avoid ethical theorising means that we refrain from merging various concerns into overarching categories. This renders the table in Annex A rather long and filled with functional overlap between the individual ethical considerations. However, the overall approach is methodologically straightforward, easily adjustable to changes in ethical standards and pragmatic in multiple ways: The ethical concerns it works with are directly relatable to prescriptive guidelines and, when combined with a data life cycle approach, it outlines a great number of potential points of contention relevant for the 'ethical case' for or against a practice (some of which may be difficult to identify in a more holistic approach) and provides a basis for coherently contesting big data practices that are structurally similar.

2 Cases

In the remainder of this article we draw on five different big data use cases from the legal domain. We select those cases primarily to cover a range of applications in the legal process, target users, target problems, system managers and countries of implementation. We summarise these relevant features of our cases in Table 1 for the five cases we select: A decision-support system for judges (COMPAS), a predictive policing system (CAS), a crowdsourcing system for lawmaking (vTaiwan) and two cases of welfare fraud detection systems (US welfare fraud detection and SyRI, which is a similar case from the Netherlands). We include the two fraud detection cases to also capture a degree of similarity and illustrate that even if two systems share similarities, they will not necessarily raise the same ethical considerations.

We now proceed case by case in the following five subsections (Sections 2.1-2.5), each offering a brief description of the case, our summary of what ethical concerns at what big data life cycle stages we can identify for that given case, and a narrative description of what makes these stages ethically consequential. That said, our ambition here is primarily exploratory, and our analysis, although systematic, is by no means exhaustive. We do not cover all six life-cycle stages for each case, primarily because we lack perfect information about those cases: for some life-cycle stages of some cases the available information is very scant, and arguing for specific ethical concerns for those stages would be over-interpreting

17. A. Applbaum, *Ethics for Adversaries: The Morality of Roles in Public and Professional Life* (2000); D. Markovits, *A Modern Legal Ethics: Adversary Advocacy in a Democratic Age* (2010).

Table 1 Characteristics of selected cases

	COMPAS	CAS	vTaiwan	US welfare fraud detection	SyRI
Legal process	Adjudication	Enforcement	Lawmaking	Enforcement	Enforcement
Target users	Judges making sentencing and bail decisions	Police officers on patrol	Legislators proposing regulation	Institutions providing welfare benefits	Institutions providing welfare benefits
Target problem	Lack of information and potential bias of judges	'Excessive' occurrence of specific crimes	Lack of public participation in lawmaking	Welfare fraud	Welfare fraud
System managements and ownership	Privately owned (Equivant)	Dutch national police	Civic community of citizens (g0v - gov zero)	Multiple public institutions	the Ministry of Social Affairs and Employment
Country of implementation	United States of America	Netherlands	Taiwan	United States of America	Netherlands

the available information and engaging in arguments for ethical positions that can be subject to reasonable disagreement, which does not align with our normatively non-committal account of ethical concerns. Furthermore, we do not want to assume that there have to be ethical concerns at every stage of the data life cycle, but we also cannot claim an absence of ethical concerns simply for want of evidence. Which stages are addressed and which are omitted is not an a priori decision (all stages are considered for each case), but for some stages we cannot argue for an ethical concern (without engaging in hypotheticals). Given the exploratory nature of this article, we choose to omit from our analysis some of these stages for some cases rather than convey a false sense of exhaustiveness that cannot be supported with the available information. This can be problematic for arriving at generalised conclusions in an article such as this, but it is less problematic in practical application: Systematising ethical concerns in this way can be done continuously, and the assessment of various stages can be 'filled in' as sufficient information becomes available. In fact, this approach can aid in identifying knowledge gaps about a case that needs to be filled in order to conduct an exhaustive analysis.

2.1 COMPAS

COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) is a need-assessment and risk-assessment tool used in the US as decision support for judges in bail and sentencing decisions. It provides an assessment of individuals' criminogenic needs that aims to aid with case planning and an assessment of risk,[18] the latter being the focus of this article. The risk assessment consists of three measures of risk calculated for a defendant – pre-trial release risk, general recidivism risk and violent recidivism risk. The purpose of these risk scores is to predict recidivism: 'The purpose of the risk scales is prediction – the ability to discriminate between offenders who will and will not recidivate.'[19] The models used to arrive at these scores are not publicly disclosed as COMPAS is a commercial product owned by Equivant (formerly known as Northpointe), and disclosing details of the system would be against their commercial interests.

The data used by COMPAS to generate the risk scores is collected by the institutions utilising it and combined with publicly available data. The data can be collected using a self-reported questionnaire or by conducting an interview during which answers to these questions are recorded.[20] This survey has 137 questions, whose answers are fed into COMPAS. Since the nature of the model utilised by COMPAS is unclear, many researchers have since studied the outcomes of COMPAS utilising data about more than 7,000 defendants in a county of Florida, including demographic details of defendants, the risk scores assigned to them by COMPAS and whether they eventually reoffend. In the case of these defendants COMPAS was approximately 65% accurate.[21] The scores themselves have been approximated by various surrogate models, but the same accuracy can also be achieved by models as simple as logistic regression utilising only two variables – age and number of prior offences.[22] Non-expert human annotators are slightly less accurate than COMPAS (62.8%) individually but more accurate than COMPAS when aggregat-

18. Equivant, 'Practitioner's Guide to COMPAS Core', www.equivant.com/practitioners-guide-to-compas-core (2019).

19. *Ibid.*, at 7.

20. Northpointe, 'COMPAS Risk & Need Assessment System Selected Questions Posed by Inquiring Agencies Ease of Use', www.northpointeinc.com/files/technical_documents/Selected_Compas_Questions_Posed_by_Inquiring_Agencies.pdf (2010).

21. J. Angwin, J. Larson, S. Mattu, and L. Kirchner, 'Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks', www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (2016).

22. J. Dressel & H. Farid, 'The Accuracy, Fairness, and Limits of Predicting Recidivism', 4(1) *Science Advances* 1, at 3 (2018).

Table 2 Ethical concerns with COMPAS

Data collection	Data acquisition & management	Analysis	Communicating insight	Decision-making	Monitoring& learning
		Independence from bias, transparency, accountability	Independence from bias, fairness, causing adverse effects for individuals	Accountability, fairness, transparency, independence from bias, respect	Independence from bias, causing adverse effects for individuals, obeying the law

ing multiple annotators together (67%). In a review of multiple instruments predicting recidivism risk in the US (including COMPAS), the authors state that 'no one instrument stood out as producing more accurate assessments than the others, with validity varying with the indicator reported'.[23]

For this case we highlight ethical concerns in four stages of the data life cycle – analysis, communicating insight, decision-making, and monitoring and learning – focusing on 'independence from bias' at every stage and illustrating the various forms of this concern, especially as related to 'accountability' and 'transparency'. These stages and the concerns we address in them are summarised in Table 2.

2.1.1 Analysis

In terms of analysis, COMPAS has been critiqued as a racially biased tool and has been shown to exhibit racial bias in the errors of the model's predictions: 44.9% of blacks labelled as 'higher risk' did not actually reoffend compared with 23.5% whites. 28% of blacks labelled lower risk did reoffend, compared with 47.7% whites.[24] Even though a defendant's race is not a feature provided to the model, other features associated with race are enough for race to arguably constitute a latent feature. This can be considered ethically significant in and of itself for the model's **independence from bias**, but the question of racial bias in COMPAS is not as straightforward and points to another ethically significant aspect – who interprets issues of justice and fairness and how it gets done: the allegation of racial bias itself is not the focus here as Equivant[25] as well as academic researchers[26] have issued rebuttals exposing serious methodological errors in the critique. What we focus on here is the ensuing debate, which, despite its largely technical character, exposed a fundamental disagreement about the meaning of 'fairness' and how it can be operationalised mathematically. Fairness can refer to accurate calibration between groups (a risk score translates to identical recidivism rate across population subgroups) or to a correct balancing of the negative and the positive classes (average risk scores for reoffenders are identical across population subgroups).[27] In other words '[t]here is no single mathematical definition of fairness. The people developing a "fair" algorithm must decide on the uniformity or variation that is necessary for a functioning system'.[28] Furthermore, this issue cannot be resolved by adjusting the algorithm, as the definition of fairness adopted by the critics and the one adopted by Equivant cannot mathematically be satisfied simultaneously unless our predictions are flawless or the base-rate of the predicted variable (reoffending) is identical for different population subgroups.[29] This means that in designing such a system one has to make the choice about what 'fairness' means (mathematically), which in this case is a decision made by technical experts without any political **accountability**, one hidden in a completely **non-transparent** algorithmic 'black box' and one that is of crucial ethical significance.

2.1.2 Communicating Insight

There are potentially ethically significant features of how COMPAS scores get communicated to judges (and other stakeholders in the legal process). Both recidivism risk scores are communicated as the score itself accompanied by a label of low risk (1-4), medium risk (5-7) or high risk (8-10).[30] The scores themselves are interpretable only with reference to a norm group and represent a specific decile of the scores of everyone in the norm group ranked in ascending order (e.g. score 1 refers to the least likely to recidivate 10% in the norm group). This means that individuals can be assigned a high risk score but not actually be highly likely to reoffend (if their norm group is generally unlikely to recidivate) and

23. S.L. Desmarais, K.L. Johnson, and J.P. Singh, 'Performance of Recidivism Risk Assessment Instruments in U.S. Correctional Settings', 13(3) *Psychological Services* 206, at 213 (2016).
24. Angwin et al., above n. 21.
25. W. Dieterich, C. Mendoza, and M.T. Brennan, 'COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity', http://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_070616.pdf (2016).
26. A. Flores, K. Bechtel, and C. Lowenkamp, 'False Positives, False Negatives, and False Analyses: A Rejoinder to "Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks"', 80(2) *Federal Probation* 38 (2016).
27. A. Chouldechova, 'Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments', 5(2) *Big Data* 1, at 2-3 (2017); J. Kleinberg, S. Mullainathan, & M. Raghavan, 'Inherent Trade-Offs in the Fair Determination of Risk Scores', https://arxiv.org/pdf/1609.05807.pdf (2016).
28. A.L. Washington, 'How to Argue with an Algorithm: Lessons from the COMPAS ProPublica Debate', 17(1) *The Colorado Technology Law Journal* 131, at 151 (2019).
29. Kleinberg et al., above n. 27.
30. Equivant, above n. 18.

vice versa.[31] The ethical significance of this is dependent on judges' understanding of how to accurately interpret these scores, but the fact that a norm group conditions the actual score a defendant receives is a potential issue for **independence from bias** and **fairness**. Furthermore, the fact that these scores are interpretable as decile scores brings into question the labelling of low, medium or high risk since judges themselves can make the decision on what particular part of the distribution of risk scores is 'high' or 'low' risk for a particular norm group and case. The system providing this label could be a source of bias in and of itself.

Another issue is communicating when exactly a risk assessment tool such as COMPAS should be utilised in the legal process. The tool itself as well as general guidelines for using risk assessments state that risk assessment 'should not be used as an aggravating or mitigating factor in determining the severity of an offender's sanction'.[32] However, the original critique levied against COMPAS mentions cases where the risk score has seemingly influenced severity of punishment when reviewed by the judge.[33] If the interpretation and use of these risk scores are not fully understood by judges it can amount to causing **adverse unjustified effects to individuals**.

2.1.3 Decision-making

COMPAS is ethically significant in three distinct ways at the decision-making stage. First, by using COMPAS the assumptions about fairness (addressed in Section 2.1.1) are made part of the legal process. Judges are the ones with the authority to interpret laws (and the conception of 'fairness' and 'justice' they capture), which appears paradoxical since they are precisely the users of the system and (knowingly or not) adopt assumptions about **fairness** made by technical experts. It is true that human judgment is often flawed and suffers from the same racial bias that COMPAS is criticised for.[34] But even if COMPAS could in some ways be less biased than judges it obscures the **accountability** for this bias: the decisions judges make and the reasoning underlying them are generally a matter of public record, and any potential bias in their decisions can be scrutinised, and they can ultimately be held accountable for it. COMPAS removes a portion of this responsibility by providing an 'impartial' and 'technical' tool whose bias is much more difficult to interrogate as the algorithm itself is a trade secret (and thus not **transparent**). Secondly, COMPAS scores are not the only piece of information judges consider when making a decision. This is related to the issue of inappropriate interpretation or overusing the tool itself (as mentioned in Section 2.1.2), but also to a more complex interaction between a risk score and other information a judge considers in a decision. For example, the socio-economic status of an individual is another important factor and one that is associated positively with risk of recidivism but negatively with the blameworthiness of an individual for the crime they have already committed. As such, providing a judge with risk assessment information can reduce the likelihood of incarceration for relatively wealthy individuals and increase this risk for relatively poor individuals (the information being identical).[35] The very inclusion of a risk assessment score can thus violate **fairness** principles and **introduce bias** to the decision-making process independently of any bias of the risk score itself.

Thirdly, introducing a tool like COMPAS into deciding individual legal cases needs to be reconciled with the individualistic nature of the legal process: any algorithm basing predictions on existing data judges individual behaviour on the basis of group characteristics. In the case of COMPAS, '[t]he moral issues involve political unease when decisions are based on immutable characteristics over which individuals have no personal control or that may serve directly or by proxy to replicate discriminatory practices'.[36] Some of the 137 features derived from a compass questionnaire are not problematic in this respect, but some are (directly or by proxy) about individuals' immutable characteristics or about their environment (e.g. criminal history of their friends and family). In general, the 'use of group tendencies as a proxy for individual characteristics'[37] is rejected in multiple pieces of US case law,[38] and the moral implications of accepting algorithmic output relying on precisely this type of inference are significant in terms of **independence from bias** and **respect** for individuals.

2.1.4 Monitoring and Learning

In terms of monitoring and learning the case of COMPAS is complicated by the lack of transparency of its inner workings, making the inspection of the algorithm itself impossible. However, the use of COMPAS can still be evaluated on the basis of its predictive outcomes and adherence to legal principles. In terms of legal principles, COMPAS has actually been legally challenged in *State vs. Loomis,* a case that ultimately reached the Wisconsin Supreme Court. In this case the defendant argued that the use of COMPAS in a sentencing decision **violates two of his legal rights** that are also ethically significant: the right to due process and the right to individualised sentence.[39] Since COMPAS is a trade secret its output cannot be scrutinised by the defence (which is a part of due process) but undeniably plays a

31. *Ibid.*
32. J. Elek, R. Warren, & P. Casey, 'Using Risk and Needs Assessment Information at Sentencing: Observations from Ten Jurisdictions', nicic.gov/using-risk-and-needs-assessment-information-sentencing-observations-ten-jurisdictions (2015), at 5.
33. Angwin et al., above n. 21.
34. Dressel & Farid, above n. 22.
35. J. Skeem, N. Scurich, and J. Monahan, 'Impact of Risk Assessment on Judges' Fairness in Sentencing Relatively Poor Defendants', 44(1) *Law and Human Behavior* 51 (2020).
36. M. Hamilton, 'Risk-Needs Assessment: Constitutional and Ethical Challenges', 52 *American Criminal Law Review* 231, at 242 (2014).
37. S.B. Starr, 'Evidence-Based Sentencing and the Scientific Rationalization of Discrimination', 66(4) *Stanford Law Review* 803, at 827 (2014).
38. *Ibid.*
39. *Case 881 N.W.2d State* v. *Loomis*, www.courts.ca.gov/documents/BTB24-2L-3.pdf (2016).

role in sentencing decisions. COMPAS also partially bases its output on aggregate data about recidivism for groups similar to the defendant but not on data for the defendant themselves (including variables like gender), potentially not delivering an individualised sentence, which would make it **biased** and cause **unjustified effects for individual defendants**. The claim of violation of due process was rejected by the court, and in commenting on the right to individualised sentence the Court upheld the use of COMPAS in sentencing decisions because it is not the determining factor for a sentence and 'is helpful in providing the sentencing court with as much information as possible in order to arrive at an individualised sentence'.[40]

In terms of monitoring the predictive outcomes of COMPAS, it is certainly possible (as evidenced by the criticism of its racial bias and the ensuing response), but there is no evidence indicating that such monitoring, which is identified as one of the guiding principles for using risk assessment tools,[41] is being done by the institutions utilising COMPAS. In terms of learning and adjustment the Wisconsin Supreme Court issued a cautionary statement to lower courts to be aware of the limitations of the COMPAS tool and the fact that it predicts group behaviour and not individual behaviour, meaning that judges need to explain factors other than the risk score that ultimately determine their decision in order to avoid undue bias and unfairness. In sum, the monitoring of the performance and legality of COMPAS was somewhat extensive, but not initiated by the institutions utilising it, and the adjustment stemming from this monitoring is very limited.

2.2 CAS

CAS (Crime Anticipation System) is a data-mining software used by most of the police forces in the Amsterdam area, the Netherlands. The software was piloted in 2014 and is currently managed by the Dutch National Police. The officially stated aim of the software is to predict the location and time of 'high impact crimes' (HIC). HIC are narrowly defined to include four offence categories, namely robbery, nuisance by youth, street robbery and bicycle and scooter theft. The software takes the form of a map where different crime categories appear as coloured squares (red for burglary, green for street robbery, blue for youth nuisance and pink for bicycle and scooter theft). Brighter intensities of the same colour indicate higher risk levels for incidents within each category (*e.g.* brighter red means a higher risk of burglary in the designated area), and each square corresponds to an area of 125 × 125 metres within the city. The combined data sources provide 78 data points for each of these squares, and the city is divided into a total of 11,500 squares. The model itself relies on a neural network, i.e. an algorithm that gradually learns to recognise and update its recognition of patterns in the data that it receives.[42]

Crime risk levels are calculated on the basis of three data sources linked together in this system. The first database is provided by the BVI (*Central Crime Database*) and includes the distance to the address of the suspect of an incident registered within the previous 6 months, the number of suspects of incidents registered within the previous 6 months that live within 500 metres, and the number of suspects of incidents registered within the previous 6 months that live within 1 kilometre from that area. The second database is provided by the CBS (*Central Statistics Office*) and currently includes 15 indicators such as the number of inhabitants, their gender, the number and average size of households, the average property and average income level, as well as the number of social benefits recipients within an area.[43] The data provided by the CBS used to include an indicator eliminated in 2017, namely the number of 'non-Western allochthones' living within an area. The third database is the BRP (*Municipal Administration*), which is used to identify streets and specific addresses (for instance, the address of a shop that might be the target of a burglary).

For this case we highlight ethical concerns in four stages of the data life cycle – data collection, data acquisition and management, communicating insight and decision-making. The key concerns appearing throughout these stages are related to 'independence from bias' and 'causing adverse effects for groups', but they do vary from stage to stage and connect to other ethical principles like 'fairness', 'respect' or 'transparency' and 'proportionality'. These stages and the concerns we address in them are summarised in Table 3.

2.2.1 Data Collection

At the stage of data collection, CAS relies on data from multiple agencies, the most ethically significant being data from the BVI. This data relies exclusively on information about the addresses of individuals who are considered crime suspects, raising a series of moral concerns. First, there are reasons to think that the list of 'suspects' might be overinclusive (and thus unreliable) because of the ethnic and classist **biases** that influence who tends to be identified as a suspect for any specific crime incident.[44] The influence of these biases might be accentuated by the fact that identifying someone as a suspect requires a relatively low evidentiary threshold and, as a result, meets little to no epistemic constraints.[45] The influence of these biases can also be com-

40. *Ibid.*, at 764.
41. Elek et al., above n. 32.
42. P. Mutsaers & N. Tom, 'Predictively Policed: The Dutch CAS Case and Its Forerunners', www.researchgate.net/publication/346593158_Predictively_policed_The_Dutch_CAS_case_and_its_forerunners/citation/download (2020).
43. S. Oosterloo & G. van Schie, 'The Politics and Biases of the "Crime Anticipation System" of the Dutch Police', in *Proceedings of the International Workshop on Bias in Information, Algorithms, and Systems 2103* (2018).
44. S. Çankaya, *De controle van marsmannetjes en ander schorriemorrie: het beslissingsproces tijdens proactief politiewerk* (2012).
45. A. Das & M. Schuilenburg, '"Garbage In, Garbage Out": Over Predictive Policing and Vuile Data', 47(3) *Beleid en Maatschappij* 254 (2020).

Table 3 Ethical concerns with CAS

Data collection	Data acquisition & management	Analysis	Communicating insight	Decision-making	Monitoring & learning
Independence from bias, causing adverse effects for groups, fairness, respect, honesty, integrity	Consent, fairness, causing adverse effects for groups, respect, transparency, data minimisation		Transparency, explainability	Causing adverse effects for groups, fairness, proportionality, respect, professionalism	

pounded (and arguably entrenched) further down the road: because of the socio-economic biases that might be incorporated into CAS *via* BVI data, police officers who follow the software's advice might gradually be brought to over-control certain areas and categories of the population and under-control others and thus gradually form or confirm a distorted image about typical offence suspects **causing adverse effects for those groups**. Furthermore, when such stereotypes about suspects influence the outputs of the system, the moral costs that come with data collection – for instance, being surveilled, stopped and interrogated – are also spread in unequal ways, which is ethically significant with regard to **fairness** and causing adverse effects to groups. Furthermore, whenever stereotypes work implicitly, the unfairly distributed costs of coping with police interference remain largely hidden in the inner workings of an algorithm.

Secondly, there is a more diffuse moral concern about the type of information that is deemed relevant for predicting future offences. By focusing on persons who are suspects in past crimes as proxies for the kinds of persons who might commit similar crimes in the future, CAS might be perceived as promoting that idea that 'no one is a suspect innocently', and thus undermine the 'innocent until proven guilty' rule that is constitutive of fair criminal justice practices. Relatedly, it might promote an objectionably stigmatising image of those thus selected as unredeemable 'criminals' or 'villains'. This militates against basic respect and equal treatment norms that should inform both police activities and the research practices on which these activities rely.

Thirdly, how CAS is presented is arguably at odds with the principles of **honesty and integrity,** given that there is a mismatch between its public image as a fine-grained predictive tool and the accuracy of the data it works with. For instance, since many burglaries and thefts happen in the absence of the victim, police officers cannot estimate the exact moment when they were committed. To deal with this problem, police officers usually choose a mid-point between the moment the victim left the house or parked the stolen bicycle and the moment when the burglary or theft was noticed.[46] The inaccurate nature of these estimates might mean that CAS is ultimately a very rough tool when it comes to calculating the timing of certain offences. Presenting it as a fine-grained tool might be empirically dishonest, at least until its effectiveness can be transparently shown.

2.2.2 Data Acquisition and Management

The data acquisition practices involved in setting up and using CAS are ethically significant in at least four distinct ways. First, it is not clear whether the acquisition of information from the CBS – in particular, data that makes it possible to geographically locate individuals living in non-traditional family settings or who are social benefits recipients – is submitted to the **consent** or oversight of the relevant human rights protection organisations.

Second, when CAS acquires data about the spatial distribution of socio-economic disadvantage as well as about the location of 'non-Western allochthones' (a practice that was terminated in 2017), for the specific goal of predicting crime, there is a risk of **unfairly** reinforcing existing stigma. The ethical significance of including this data is emphasised because of the absence of conclusive evidence that socio-economic disadvantage or ethnic difference are causally linked to higher crime rates.[47] The very inclusion of such data shows a willingness to single out individuals with a specific social and ethnic background as potential criminals. This can constitute a violation of basic norms of **respect** and principles of equal treatment and not **causing unjustified adverse effects for groups or individuals**.

Third, the Dutch National Police provide no publicly accessible information about the list of indicators and data they acquire from other sources, such as the CBS or BRP. This is arguably an infringement of **transparency**, as it leaves citizens in the dark about the considerations that guide police surveillance and other forms of interference that affect their and their co-citizens' lives. One reply here might be that the precise indicators and data included in CAS cannot be publicly advertised because doing so would affect its predictive effectiveness – for instance, by allowing some future offenders to foresee where and when police forces will be deployed. This rejoinder, however, does little to alleviate transparency concerns, especially as we currently lack evidence about the effectiveness of CAS.

46. Oosterloo & van Schie, above n.43.

47. T. Newburn, 'Social Disadvantage: Crime and Punishment', in D. Hartley and L. Platt (eds.), *Social Advantage and Disadvantage* (2016).

Fourth, CAS raises concerns about **data minimisation** requirements: when the decision was made in 2017 to exclude information about the number of 'non-Western allochthones' living within an area, the administrators of the software argued that the variable did not add to its predictive power.[48] This means that, for approximately 3 years following its introduction, CAS was substantively violating **data minimisation** requirements. This decision raises a more general question about the extent to which the specific indicators included in CAS are needed to ensure that it performs well in predicting the timing and location of crime incidents.

2.2.3 Communicating Insight

In terms of communicating insight, CAS is a closed system,[49] which means that users have access to information only about the risk levels of a particular crime category in a given area based on a fixed number of variables. Consequently, police officers do not have the option of zooming in on an area that is designated as high risk for a crime category (e.g. street robbery) to get more information that would allow them to make more empirically informed hypotheses about the context or seriousness of the crime risk level going upwards in that particular area within a specific time interval. This limitation of CAS can be ethically significant for the principle of explainability, if and insofar as this principle applied reflexively to the police officers themselves, and not only (as is usually the case) to the citizens who are policed. In using CAS, police officers might not be able to understand why a particular output was reached by its underlying algorithm, and, by way of consequence, they might not be able to explain why the output was reached to those they police. Also, the fact that CAS remains insensitive to whether an increase in crime risks is driven by any particular stable variable included in the software or by more conjunctural events (e.g. street parties) is significant for **transparency** as it does not allow users to properly grasp why crime patterns are changing in any particular area at a given moment in time.

2.2.4 Decision-making

CAS raises at least two distinct moral problems when it comes to policing decisions that are based on it. First, and insofar as it relies on 'dirty data'[50] that carries forward patterns of discrimination and disadvantage, CAS-based policing could contribute to compounding or entrenching the disproportionate amount of surveillance and interference that some neighbourhoods and categories of the population are submitted to, and thus **causing unjustified negative effects for groups**. This could violate both **fairness** principles (by upsetting the fair distribution of social burdens across persons and groups within the general population) and **proportionality** principles (with the benefits generated by the application of CAS largely unknown, it is difficult to determine whether the risks that come with over-policing are justified). Second, as police officers have noted themselves,[51] focusing too much on the advice given by CAS reduces policing to prevention and thus diverts officers from other obligations they are expected to tend to, such as assisting people with the coordination of various social activities or establishing a rapport with the inhabitants of any particular neighbourhood. At this level, the focus that CAS puts on prediction and prevention might be in tension with the **professionalism and respect** that citizens can also reasonably expect from their police officers.

2.3 vTaiwan

vTaiwan (v stands for vision, voice, vote and virtual) is a legislative crowdsourcing system used since 2015 by the Taiwanese government to give citizens a way to propose and debate new laws with the output of this discussion ideally influencing future legislation. Conceptually, legislative or policy crowdsourcing 'involves giving ordinary citizens, rather than political and bureaucratic elites, the chance to cooperate to come up with innovative new policies'.[52] It can thus be used for both policymaking and statutory lawmaking. In the latter form, it involves collaborative lawmaking between official lawmakers and networks of citizens and civil society organisations that aims to build the quality of legislative documents and increase political legitimacy of new legislation.[53] It also involves a new kind of role for citizens in the legislative process, moving from top-down models of legislative development to approaches that address information asymmetries between professionals and consumers or citizens, giving the latter more influence.[54] The data for the vTaiwan system is contributed by citizens in a range of different forms such as social media comments, discussion forums or online petitions. These contributions are then analysed using big data analysis instead of the traditional approach of being comprehended only through time-intensive human reading of texts.

Many countries use online platforms as a resource for members of the public to track laws proposed by parliament and make comments about them (e.g. regulations.gov in the US or Avoin Ministerio in Finland), and there have also been notable crowdsourcing approaches to specific legislative initiatives such as the

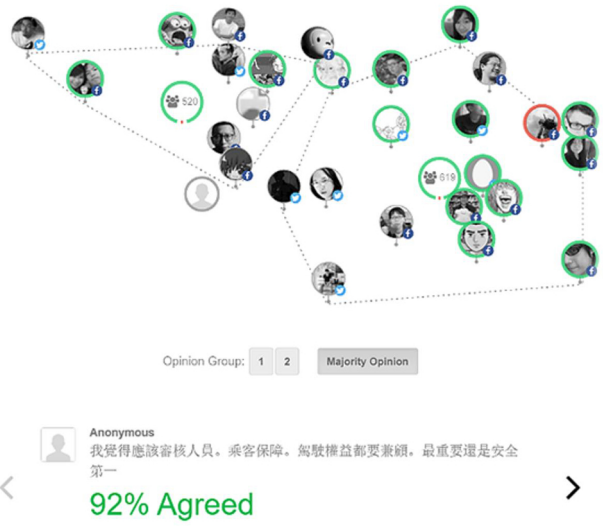
48. Oosterloo & van Schie, above n. 43.
49. *Ibid.*
50. Das & Schuilenburg, above n. 45; P. Mutsaers, 'A Public Anthropology of Policing. Law Enforcement and Migrants in the Netherlands' (dissertation at University of Tilburg) (2013).
51. A. Drenth & R. van Steden, 'Ervaringen van straatagenten met het Criminaliteits Anticipatie Systeem', 79(3) *Het tijdschrift voor de Politie* 6 (2017).
52. H.S. Christensen, M. Karjalainen, and L. Nurminen, 'Does Crowdsourcing Legislation Increase Political Legitimacy? The Case of Avoin Ministeriö in Finland', 7(1) *Policy & Internet* 25 (2015).
53. V. Burov, E. Patarakin, and B. Yarmakhov, 'A Crowdsourcing Model for Public Consultations on Draft Laws', in *Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance* (2012).
54. T. Heikka, 'The Rise of the Mediating Citizen: Time, Space, and Citizenship in the Crowdsourcing of Finnish Legislation', 7(3) *Policy & Internet* 286, at 287 (2015); S. Ranchordás and W. Voermans, 'Crowdsourcing Legislation: New Ways of Engaging the Public', 5(1) *The Theory and Practice of Legislation* (2017).

Table 4 Ethical concerns with vTaiwan

Data collection	Data acquisition & management	Analysis	Communicating insight	Decision-making	Monitoring & learning
		Respect, fairness, independence from bias, transparency, explainability	Honesty, independence from bias	Accountability, honesty, respect, principled performance	

Internet Bill of Rights in Brazil[55] or the new constitutional processes in Ireland and Iceland.[56] vTaiwan is a notable case because of the original way that it analyses and openly presents digital information about its users in real time to facilitate deliberation. It has four main stages in the development of legislation: proposal, opinion, reflection and approval. In the opinion stage stakeholders are identified, and more stakeholders are included using a rolling survey, followed by gathering and analysing a large number of public opinions with the goal of distinguishing important clusters of topics that can be visualised in the form of network diagrams. These models have the explicit purpose of facilitating consensus: users can up-vote or down-vote suggestions (but cannot comment on them) or issue their own suggestions, and the algorithm automatically separates those suggestions into 'opinion groups' and de-emphasises the areas of disagreement between them. This system results in people looking for consensus across various 'opinion groups' and creating new suggestions that even people from disparate 'opinion groups' will upvote. An example of such a network diagram is provided in Figure 2. Once a satisfactory level of consensus has been reached a given round of opinion mining is concluded. The algorithm itself utilises the pol.is open-source system, which relies mainly on principal components analysis (PCA) and k-means clustering to obtain and visualise 'opinion groups'.[57]

Figure 2 Example of opinion groups in a network diagram[58]



In this case we focus on ethical considerations in three stages of the data life cycle – analysis, communicating insight, and decision-making – mainly highlighting the role of 'honest' and 'fair' summarisation of citizens' opinions but also touching on how this relates to 'independence from bias', 'explainability' or 'principled performance' of the system itself. These stages and the concerns we address in them are summarised in Table 4.

2.3.1 Analysis

In the analysis stage vTaiwan is ethically significant in terms of respect, fairness and the potential for containing biases. The analysis of opinion data is predicated on the programmers' understanding of how 'opinion groups' should be constructed and what consensus (or lack thereof) looks like between these groups. The task of accurately modelling a corpus of texts is in and of itself reliant on subjective assumptions about what content is relevant and how it should be described, which is in this case amplified by reducing this information down to a two-dimensional space to be able to visualise the clustering that defines 'opinion groups'. The two principal components defining this space do not have an

55. D. Arnaudo, 'Computational Propaganda in Brazil: Social Bots During Elections', *University of Oxford Working Paper* 8 (2017).
56. S. Suteu, 'Constitutional Conventions in the Digital Era: Lessons from Iceland and Ireland', 38 *Boston College International and Comparative Law Review* 251 (2015).
57. Y.T. Hsiao, S.Y. Lin, A. Tang, D. Narayanan, and C. Sarahe, 'vTaiwan: An Empirical Study of Open Consultation Process in Taiwan', https://doi.org/10.31235/osf.io/xyhft (2018).
58. Figure reproduced from NESTA, 'vTaiwan' www.nesta.org.uk/feature/six-pioneers-digital-democracy/vtaiwan/ (last visited 16 November 2020).

58. Figure reproduced from NESTA, 'vTaiwan' www.nesta.org.uk/feature/six-pioneers-digital-democracy/vtaiwan/ (last visited 16 November 2020).

inherent meaning and can be constructed in various ways, none of which are strictly 'wrong' or 'right'. This means that how opinions are grouped, what is considered salient for legislative formation and what is consensus between those groups can be expressed in multiple ways. By selecting one of those ways some opinions inevitably get downplayed, others get up-played, and some opinions that do not fit well into large 'opinion groups' might effectively be silenced by being aggregated into these groups (de-emphasising their uniqueness). Summarising opinions in this way can be violative of norms of **respect** for individuals (and their opinion) and **fairness**, and can also introduce a **bias** in terms of what gets highlighted and what gets lost.

The concern with bias is also relevant because crowdsourced legislative commentary is highly diverse in terms of the kinds of populations that may be involved. Crowdsourcing social media data for legislative development can sometimes circumvent this problem if the data is representative of a population at large. However, the problem is most acute if the analysis is of crowdsourced commentary that may be contributed disproportionately by specific interest or demographic groups. This **bias** emerges from the moment the data is collected, but it also affects the roles that different kinds of citizens play in the monitoring of analysis. Even in a technologically advanced country such as Taiwan, digital skills and access inequalities exist among population subgroups such as the elderly or less wealthy. Achieving fairness in such circumstances is vital, but more than being a pervading ethical principle, it must also have safeguards provided by measures to ensure transparency and explainability of algorithms used in the analysis as well as information about the representativeness of those who contribute their opinions. vTaiwan does particularly well with regard to **transparency** principles as it is based on several open source platforms, but **explainability** is much more challenging as not only is computer code understood only by a small section of the population but the decisions in model specifications need to be interrogated in terms of their impact and justified more than simply being made transparent.

2.3.2 Communicating Insight

In terms of communicating insight, many of the concerns previously addressed apply here as well. In some ways, the analysis in this case is, at its core, a communication exercise: the aim of the model itself is to provide a summary of the crowdsourced opinions in a way that is understandable and conducive to consensus seeking. This is particularly problematic given how the delivery of crowdsourced legislation attenuates conflicting interests of analysts and politicians where the former are focused on the technical quality of analysis and the latter are tasked with turning the results of the analysis into actionable results with political consequences. In this case vTaiwan has not yet been used for major legislation that would make those types of challenges sufficiently apparent, but even if its performance is good the ethical significance of visualising an extremely multifaceted data (written opinions) in a two-dimensional space without introducing **bias** remains.

The communicating insight phase is crucially important for this case because democratic systems that rely on citizen input are implicitly (and often explicitly) responsible for reporting back to citizens on what resulted from their contributions. In this respect, communicating insights should, in its general form and processes, be an honest and accurate interpretation of the crowdsourced data, which is very difficult to assess in this case as different analysts and politicians might genuinely see different summaries of the texts as 'accurate'. The **honesty** concern here is particularly pertinent, since the platform aims to de-emphasise disagreement and foster consensus, arguably not summarising the corpus of texts in a fully 'honest' way. The attempt to communicate with citizens in this way should also show democratic values such as equality and understanding of the contextual situation of citizens by making the insights understandable to citizens of different education and technical levels. For citizens with more technical skills, open access to the data and programming steps needed to reproduce the analysis are essential for supporting principles of equality as well as transparency and accountability.

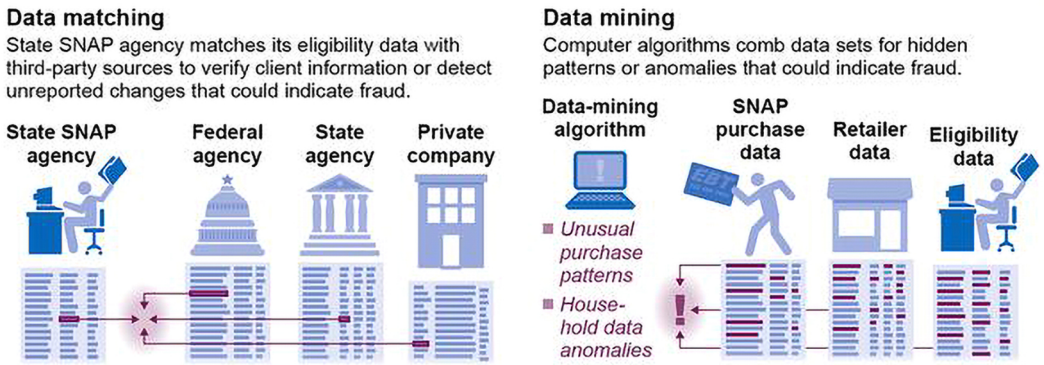
2.3.3 Decision-making

The decision-making stage of this case is particularly contentious as further political steps such as parliamentary debate and voting are required for citizens' opinions to have any effect. This stage actually substantially affects the ethical concerns that are relevant in previous stages: even though the system is generally transparent and open to the public, the government does not have to follow its outputs, making its **principled performance** uncertain. This conjures concerns of 'open washing' by having a system that is transparent and provides the government with legitimacy in decision-making but one that can ultimately be easily ignored when it comes to legislation.[59] This is a concern relevant to **respect** for citizens' opinions and **honesty** but also one that impacts the entire life cycle – if there is no **accountability** of decision makers for simply dismissing the output of this system, the system should be viewed in a different light.

Depending on where the decision-making process falls on the technical-political spectrum, the ethical principles will be different. For technical decisions, transparency of the technical decision-making processes involved in turning raw data into new legislative changes is of primary importance. For political decisions, legal and constitutional compliance in the ways that the political decision-making process are followed are of primary importance. Both technical and political decisions share the ethical problem of respecting citizen rights and dignity and protecting citizens from adverse effects that may result from changes to the law.

59. *Ibid.*

*Figure 3 Data matching and mining**



* Figure retrieved from GAO, above n. 60, at 1.

2.4 US Welfare Fraud Detection

Data surveillance to detect various types of fraud in US welfare programmes is an overarching system of data collection, management, analysis, and decision-making with the purpose of discovering fraud in utilisation of welfare programmes. Exact algorithms to identify fraudulent behaviour vary and are not fully transparent, but they are generally scouring large linked databases to identify patterns indicating potential fraud. These databases are constructed by mining and matching data across program-specific databases such as the Supplemental Nutrition Assistance Program (SNAP), federal and state agencies as well as private company data. This cross-referencing is particularly relevant to the SNAP programme, because it allows recipients to spend their benefits outside their state of residence.[60] Figure 3 demonstrates the extent of the database linking efforts spanning state agencies, federal agencies as well as private companies. Most of this data collection is done at the state level, and states have discretion in how to handle the information. In addition, 'current legal frameworks offer little protection for privacy-related harms experienced by the poor', giving additional leeway to the government for utilising such data.[61]

Figure 3 also touches on the fact that in the era of big data these efforts have gained even more momentum. One example is the replacement of food stamps with the Electronic Benefit Transfer (EBT) Card. The prepaid debit card provides an electronic way to pay in stores without showing stamps but also gives government the opportunity to track purchases. Another new data source being integrated is social media: some services require the use of specific platforms by potential welfare recipients in order to access information and resources. Currently, the US government is planning to use social media profiles to detect welfare fraud with some monitoring already utilised in fraud and abuse investigations.

> In 2014, the SSA's Office of the Inspector General (OIG) utilized social media reviews to help arrest more than 100 people who defrauded Social Security Disability Insurance (SSDI) out of millions of dollars. Investigators found photos on the personal accounts of disability claimants riding on jet skis, performing physical stunts in karate studios and driving motorcycles.[62]

In this case we focus on ethical concerns in four stages of the big data life cycle – data collection, data acquisition & management, decision-making and monitoring and learning. The primary ethical concern here has to do with 'bias' and the resulting undue 'adverse effects for individuals and groups', touching on some additional unequally distributed 'privacy' concerns as well as 'principled performance' concerns raised by automating systems like this. These stages and the concerns we address in them are summarised in Table 5.

2.4.1 Data Collection

The primary ethical concern at the stage of data collection is the over-surveillance of certain population subgroups, in this case welfare recipients. In the US, low-income individuals are more likely than others to experience monitoring by the government, which is relevant to **causing adverse effects to groups**. In fact, 'low-income communities are among the most surveilled communities in America'.[63] This goes back to the 1996 welfare reform bill entitled 'Personal Responsibility and Work Opportunity Reconciliation Act' (PRWORA), which calls for an elaborate system of performance indi-

60. GAO, 'Supplemental Nutrition Assistance Program: Disseminating Information on Successful Use of Data Analytics Could Help States Manage Fraud Risks', www.gao.gov/products/GAO-19-115 (2018).
61. M. Madden, M. Gilman, K. Levy, and A. Marwick, 'Privacy, Poverty, and Big Data: Matrix of Vulnerabilities for Poor Americans', 95(1) *Washington University Law Review* 53, at 113 (2017).
62. M. Miller, 'U.S. Government Weighs Social-Media Snooping to Detect Social Security Fraud', www.reuters.com/article/us-column-miller-socialmedia-idUSKCN1RA12R (2019), at 2.
63. K. Waddell, 'How Big Data Harms Poor Communities', www.theatlantic.com/technology/archive/2016/04/how-big-data-harms-poor-communities/477423/ (2016), at 1.

Table 5 Ethical concerns with US welfare fraud detection

Data collection	Data acquisition & management	Analysis	Communicating insight	Decision-making	Monitoring & learning
Privacy, independence from bias, causing adverse effects for groups	Independence from bias, causing adverse effects for groups, privacy, fairness			Independence from bias, causing adverse effects for groups, principled performance	Accountability, independence from bias, causing adverse effects for groups, transparency

cators with the main goal of 'welfare-to-work' efforts.[64] The establishment of those indicators allows intrusive data collection. For example, it 'empowered state governments to delve into the personal and sexual lives of women of all ages, by requiring single mothers to identify the biological fathers of their offspring' and by capping welfare payments if women had more children while in the programme.[65] Collecting this data is done on top of cross-referencing databases, following up on tips from welfare fraud hotlines, drug-testing and physically surveilling the poor.[66] Digitisation of certain provisions such as the adoption of EBT cards further highlights this issue, as electronic EBT transactions data shows purchase histories and amounts that were spent. Already in 1999, there was a discussion around **privacy** intrusion based on the digitising services, such as EBT, that would apply only to those in need of government support.

> Previously recipients could anonymously cash their checks or spend their food stamps. That is, the transaction did not link the individual to the purchase. With EBT, a permanent record of precisely what the person does with the government benefit often will be created.[67]

Beyond data that is collected at the individual level, there is another dimension of neighbourhood surveillance. Welfare recipients tend to live in poorer neighbourhoods, which are also subjected to more police presence and CCTV monitoring. In addition, people who live in crowded, urban neighbourhoods are more likely to suffer warrantless searches by government agents.[68] 'As a result, they are much more likely than other people in other contexts to become entangled with the criminal justice and child welfare systems, both of which are highly stigmatizing and privacy-stripping'[69] – all of these interactions are ethically relevant with regard to **independence from bias**. These interactions then become embedded in large and linked databases that use this input to assess other things than fraud detection such as housing, employment or educational opportunities.[70]

2.4.2 Data Acquisition and Management

This case is ethically significant in terms of data acquisition owing to its utilisation of social media data and its ambition to extend this practice to frontline workers who work with claimants.[71] This is problematic in two ways: first, social media data makes it possible to assess a person's network as well as create a profile on the basis of online behaviour and preferences. 'Poor Americans have long suffered from guilt by association, meaning they bear the stereotypes and stigma of their social class (and race and gender) in ways that impede their economic progress and well-being',[72] which is relevant to potential violations of both **privacy** principles and **fairness** principles. And second, potential knowledge gaps around privacy can make welfare recipients' 'privacy vulnerable', in particular because of their reliance on mobile connectivity and fewer restrictions they put on the content being posted online.[73] Acquisition and utilisation of this data that has an inherent **bias** could reinforce existing patterns of neglect and socioeconomic disadvantage, resulting in **adverse effects for groups and individuals**. The fact that these conclusions are reached by leveraging individuals' associations within their social network is seemingly not aligned with the individualistic nature of the legal process and raises **fairness** concerns.

2.4.3 Decision-making

There are additional ethically significant concerns linked with the move towards automated decision-making in this case. Many states have started to use data-

64. N. Maréchal, 'First They Came for the Poor: Surveillance of Welfare Recipients as an Uncontested Practice', 3(3) *Media and Communication* 56 (2015); M. Wiseman, 'Welfare Reform in the United States: A Background Paper', 7(4) *Housing Policy Debate* 595 (1996).
65. Maréchal, above n. 65; B. O'Connor, *A Political History of the American Welfare System: When Ideas Have Consequences* (2003).
66. Maréchal, above n. 65.
67. P.P. Swire, 'Financial Privacy and the Theory of High-Tech Government Surveillance', 77(2) *Washington University Law Quarterly* 461, at 505-6 (1999).
68. M. Gilman, 'AI Algorithms Intended to Root Out Welfare Fraud Often End Up Punishing the Poor Instead', https://theconversation.com/ai-algorithms-intended-to-root-out-welfare-fraud-often-end-up-punishing-the-poor-instead-131625 (2020).
69. Madden et al., above n. 61, at 66.
70. *Ibid.*
71. Miller, above n. 63.
72. Madden et al., above n. 61, at 66.
73. Madden et al., above n. 61.

mining techniques for data analysis and automatic identification of fraud in, for example, the food stamp programme or unemployment insurance. In this effort, states are receiving the federal government's support to upgrade their technology and software. Recent examples show that this automation of fraud detection in combination with payouts is not reliable, raising questions about the **principled performance** of these systems:

> In Michigan, a $47 million automated fraud detection system adopted in 2013 made roughly 48,000 fraud accusations against unemployment insurance recipients – a five-fold increase from the prior system. Without any human intervention, the state demanded repayments plus interest and civil penalties of four times the alleged amount owed … As it turns out, a state review later determined that 93% of the fraud determinations were wrong.[74]

Michigan is not the only state experiencing problems with automated decision-making in fraud detection. Similar issue are reported from Indiana, Arkansas, Idaho and Oregon.[75] Reasons for such faulty systems are manifold. States lack funding, skilled analysts as well as data to make automated decision-making systems work.[76] This results in both bias and adverse effects for individuals in the decision-making stage.

2.4.4 Monitoring and Learning

In terms of monitoring these data processes, these issues in decision-making as well as in data collection and acquisition are hard to track because welfare programmes are the responsibility of the state and the federal government has little to no authority to 'oversee or assess the adequacy of benefit levels, bureaucratic process, or the return on investment in terms of assuring a decent quality of life for the poorest'.[77] There is thus little control over these automated processes that further entrench existing **bias** and result in **adverse effects for individuals** and vulnerable groups without much human oversight. This raises issues around accountability and transparency when it comes to retracing the steps that were taken and the ability to assess the process.

2.5 SyRI – Dutch Benefit Fraud Detection

SyRI (System Risk Indication) was created by the Dutch government following legislation passed by the Dutch Parliament in 2014. Multiple national governmental bodies can request the minister to use the system, including Dutch municipalities, the Employee Insurance Agency, the Social Security Bank, the Tax Authority, and the Ministry of Social Affairs and Employment. Governmental agencies contributing data to the system were even more numerous, as SyRI was originally allowed (by the legislation allowing its operation) to utilise 17 categories of government data, including taxes, fines, residence, debts and benefits. Because of this breadth of included personal data, the Council of State recommended to install a 'select before you collect' principle, as per the Council's conclusions published in the Staatscourant.[78] The principle requires that parties first determine what data is needed to achieve the objective and then only selectively acquire the data needed, rather than collecting all data accessible to them.

The totality of this data was then fed into an 'artificial intelligence algorithm', the details of which remain secret to this day. To use SyRI one of the agencies authorised to utilise it would request the Ministry of Social Affairs and Employment and identify a neighbourhood they believe to have an elevated risk of benefit fraud. SyRI can then identify specific individuals and addresses in those neighbourhoods that pose an elevated risk of benefits fraud or misuse. Any such risk identification by the system is not a form of evidence of a violation and in and of itself cannot be used for law enforcement[79] – the goal is to identify cases for further inspection and communicate those (excluding false positives the ministry itself can identify) to the agency making the request.

In this case we focus on ethical concerns in three stages of the big data life cycle – data acquisition and management, analysis and monitoring and learning. The primary ethical concerns here have to do with 'privacy', 'data minimisation', and 'proportionality' of the gathered data. These stages and the concerns we address in them are summarised in Table 6.

2.5.1 Data Acquisition and Management

The case of SyRI is ethically significant here because of the relatively unchecked breadth of data sources it links together, potentially violating the principles of **data minimisation** and appropriate balancing of invasiveness with societal benefits of a system. Even as SyRI was being established, the Council of States noted that the list of personal data that may be utilised by SyRI is 'so broad that it is hardly possible to think of any personal data that does not fall under it. The list does not seem intended to be limiting, but to have as much leeway as possible'.[80] The ethical consideration of **proportionality** applies here as well (the risks to individual privacy should be proportional to the societal benefit). In this case the threat to **privacy** is certainly substantial because of both the sheer variety and volume of data SyRI utilised and the fact that individuals simply have to live in the 'wrong' neighbourhood to be potentially analysed by SyRI. The benefit seems to be rather ques-

74. Gilman, above n. 69.
75. *Ibid.*
76. T. Newcombe, 'Aiming Analytics at Our $3.5 Billion Unemployment Insurance Problem', www.govtech.com/data/Aiming-Analytics-at-Our-35-Billion-Unemployment-Insurance-Problem.html (2017).
77. Maréchal, above n. 65, at 63.
78. Staatscourant, 'Raad van State. Ontwerpbesluit houdende regels voor fraudeaanpak door gegevensuitwisselingen en het effectief gebruik van binnen de overheid bekend zijnde gegevens (Besluit SyRI)', *Staatscourant nr. 26306* (2014).
79. Dutch Government, 'Answer to Parliament Questions 2018Z18418', *Buitenweg. Ref. 2018-0000177182* (2018).
80. Staatscourant, above n. 79 (authors' own translation).

Table 6 Ethical concerns with SyRI

Data collection	Data acquisition & management	Analysis	Communicating insight	Decision-making	Monitoring & learning
	Data minimisation, proportionality, privacy	Transparency, explainability, causing adverse effects for groups			Data minimisation, proportionality, transparency, privacy, obeying the law, causing adverse effects for groups

tionable as SyRI analysis was conducted only in four Dutch cities and likely resulted in no new cases of fraud identified.

There is also a secondary and a much more practical concern that is potentially also ethically significant in the data management of the SyRI system linked to data storage: in a reply to parliamentary questions, the government acknowledged that source files were not always destroyed when they should have been but in given cases almost 18 months too late.[81] The data was stored in a secure area that could be accessed only by those authorised to work with SyRI, but this does exacerbate the **proportionality** concerns as personal information was exposed to a greater risk of misuse or potential security breach than necessary with no demonstrable societal benefit.

2.5.2 Analysis

This case is ethically significant for analysis in two distinct ways: first, a lack of clarity in the indicators and risk profiles used may lead to a 'fishing expedition'.[82] Even though the SyRI law originally allowed the practice, the data acquired for this system was in different systems, administered by different organisations and collected for different goals. Inappropriate use of this data may eventually lead to reluctance to share it with the government, ultimately limiting effectiveness. The government opted for not disclosing data sources and methods of analyses to avoid disclosing the modus operandi and thus lowering the risks of those committing fraud and gaming the system.[83] This is a clear instance of **transparency** and **explainability** being sacrificed to maintain the effectiveness of the system (in terms of data acquisition and the difficulty of circumventing it by malicious actors).

Secondly, there are ethical concerns with regard to reinforcing existing stigmatisation and discrimination as SyRI 'benefits from a relatively clear public legal framework' despite its 'alleged discriminatory character'.[84] Especially the targeting of specific areas led to civil advocacy against the system and to accusations of discriminatory or stigmatising effects of the system owing to its use of a broad range of data likely including protected characteristics. This can result in over-surveillance of individuals based on existing stigmatising patterns or the over-surveillance of entire neighbourhoods based on factors largely beyond the control of any individual living within it, both of which are undue **adverse effects for individuals and groups**.

2.5.3 Monitoring and Learning

In the case of SyRI the monitoring and learning process was rather public. In fact, what make this case well known are reports by international news and professional media[85] related to a lawsuit that made it a test case for algorithmic governance. In this case the Court ruled that the law establishing SyRI was in violation of the European Convention on Human Rights because it was too invasive,[86] making this ethically relevant in terms of **obeying the law** but also **privacy**. In its ruling, the Court argued that deployment of new technologies towards these ends can be legitimate but also that the government has a special responsibility to find the right balance between deploying such technologies for the public good and respecting and protecting privacy, referring to **proportionality**. The ruling concluded that using this much data on this level violates private life, does not fit with principles of **transparency** and restraint in data use (**data minimisation**) and creates risks that the system might discriminate and thus cause **adverse effects for individuals and groups**,[87] leading to the immediate termination of SyRI.

3 Discussion

An overview of the case summaries presented in Section 2 is presented in Table 7, which combines the individual one-row tables used for each of the preceding

81. Dutch Government, above n. 80.
82. Staatscourant, above n. 79.
83. Dutch Government, above n. 80.
84. S. Ranchordás and Y. Schuurmans, 'Outsourcing the Welfare State: The Role of Private Actors in Welfare Fraud Investigations', 7 *European Journal of Comparative Law and Governance* 5, at 6 (2020).
85. The Economist, 'Humans Will Add to AI's Limitations', www.economist.com/technology-quarterly/2020/06/11/humans-will-add-to-ais-limitations (2020); T. Simonite, 'Europe Limits Government by Algorithm. The US, Not So Much', www.wired.com/story/europe-limits-government-algorithm-us-not-much/ (2020).
86. Rechtbank Den Haag, Ruling ECLI:NL:RBDHA:2020:865 (2020).
87. *Ibid*.

Table 7 Ethical concerns in selected cases throughout the big data life cycle

	COMPAS	CAS	vTaiwan	US welfare fraud detection	SyRI
Data collection		Independence from bias, causing adverse effects for groups, fairness, respect, honesty, integrity		Privacy, independence from bias, causing adverse effects for groups	
Data acquisition & management		Consent, fairness, causing adverse effects for groups, respect, transparency, data minimisation		Independence from bias, causing adverse effects for groups, privacy, fairness	Data minimisation, proportionality, privacy
Analysis	Independence from bias, transparency, accountability		Respect, fairness, independence from bias, transparency, explainability		Transparency, explainability, causing adverse effects for groups
Communicating insight	Independence from bias, fairness, causing adverse effects for individuals	Transparency, explainability	Honesty, independence from bias		
Decision-making	Accountability, fairness, transparency, independence from bias, respect	Causing adverse effects for groups, fairness, proportionality, respect, professionalism	Accountability, honesty, respect, principled performance	Independence from bias, causing adverse effects for groups, principled performance	
Monitoring & learning	Independence from bias, causing adverse effects for individuals, obeying the law			Accountability, independence from bias, causing adverse effects for groups, transparency,	Data minimisation, proportionality, transparency, privacy, obeying the law, causing adverse effects for groups

cases. Table 7 provides a structured summary of what we found to be important ethical considerations at various stages of the big data life cycle.

Despite the exploratory nature of this systematic overview and the largely illustrative case selection, we can make a few insightful observations. Primarily, we show that relevant ethical concerns can indeed emerge across the entire big data life cycle, substantiating the arguments that claim this to be the case.[88] This is not a surprising finding, as it is generally accepted that this is the case, but this article is innovative in that it operationalises this approach and shows that this intuition applies to the legal domain.

Table 7 can be read in multiple ways. Reading the table as a whole shows that issues of bias and adversely affecting individuals and groups are the most frequent ethical considerations. Other issues such as transparency are also prominent in the table as a whole, but some issues, such as accountability, privacy or obeying the law, can be identified far less often. This observation can be enhanced by reading the table row by row, by which means issues of bias and adversely affecting individuals or groups are shown to be cross-cutting and can be identified in every single data life cycle stage multiple times, even though we focus only on five illustrative

88. Wing, above n. 8.

cases. Other considerations such as transparency remain relatively cross-cutting but clearly over-represented in certain life cycle stages (in the case of transparency this is the 'analysis' stage). Other considerations remain far more circumscribed to a specific life cycle stage, such as obeying existing law, which we can identify only at the stage of monitoring and learning in all five cases. Needless to say, the generality of these observations is limited by analysing only five distinct cases and by the limited available information about these cases, but it is an indication that some ethical considerations tend to apply more to specific big data life cycle stages than to others.

The more interesting observation comes from reading the table column by column (case by case); this shows the interconnectedness of individual stages in any given case and allows us to get a better grasp of how this interconnectedness plays out for ethical concerns. It seems that there are situations where a key concern emerges in multiple stages in a slightly different form. This suggests that concerns from earlier stages of the data cycle can get 'transferred further' or even compounded throughout the data life cycle. The compounding is apparent in, for example, the CAS case, where data collection itself results in 'dirty data' owing to bias in suspect identification and acquisition of specific type of data about ethnicities reflects a further discriminatory assumption, culminating in concerns about discrimination at the level of decision-making. However, it also seems that it is not just a question of issues earlier in the data cycle influencing what happens next – even issues that happen later in the cycle can be significant for ethical concerns at a preceding stage. This is apparent in the case of vTaiwan, where the risk for legislators to ignore the analysis at the stage of 'decision-making' can influence what the relevant ethical concerns are earlier in the cycle. We believe this effect can also be positive, for example, anonymisation of data during the 'data management' stage can alleviate privacy concerns that took place at the 'data collection' stage. This also raises questions for future research, for example, whether bias at the data collection stage carries through to the acquisition and analysis stages or whether new or additional forms of bias are introduced at that point in the life cycle. It also facilitates a discussion around whether rectifying bias at the collection stage will solve bias-related challenges later on in the life cycle or whether they are reintroduced in a different way.

4 Conclusions and Limitations

This article adopts a process-oriented definition of big data and a relatively simple model of the big data life cycle to systematise ethical concerns along the stages of this life cycle. To do so, it adopts an ethically pluralistic and pragmatic perspective on 'ethical' concerns and selects five cases that together capture a broad range of big data uses in the legal domain: a decision-support system for sentencing and bail decisions (COMPAS), a predictive policing system (CAS), a legislative crowdsourcing system (vTaiwan) and two welfare fraud detection systems (one deployed in the US and the other in the Netherlands). Discussing each case in turn, the article provides an overview of these cases, delivering on the intended systematic summary and making a few interesting observations. In particular, the life cycle perspective is capable of highlighting how ethically significant practices and choices may manifest themselves differently in different stages of a use case.

Despite delivering the intended systematic summary, the article has a few limitations worth highlighting: first, the various ethical concerns we refer to throughout the article (and that we list in Annex A) are not mutually exclusive but considerably overlap. This is a direct result of our decision not to merge or re-categorise the identified ethical concerns, as that would necessitate distinctive normative commitments. As a result, some ethical concerns will often incorporate other ethical concerns, and there are multiple ways to label a given issue. For example, anything consequential for the 'independence from bias' is often also related to 'not causing unjustified or adverse effects for individuals or groups' (because that is what biased data tends to result in) or to 'respect' or 'fairness' (as that is often violated by treating individuals as stereotypical examples of a group they belong to). This limitation is important to highlight because it suggests that the number of ethical concerns we include in a stage does not necessarily reflect the ethical 'seriousness' of any particular practice. Put differently, the ethical considerations we list and highlight in the five cases do not necessarily aggregate in the balance of moral concerns. Of course, others attempting to do a similar systematisation, especially when it comes to practitioners implementing or morally interrogating a big data system, can have a more committed and normative definition for ethics to resolve this issue. Second, the stages of the big data life cycle are not as distinct in practice as they are in our model and systematisation. The fact that these stages functionally overlap is apparent from, for example, the case of vTaiwan, where the goal of analysis is to communicate something in a specific way, which makes the stages of 'analysis' and 'communicating insight' inseparable. Sometimes even clearly distinct stages are ethically intertwined. Consider the value of 'due diligence to evaluate data practices of third-party collaborators', which implies that even if data is only obtained and not collected, data collection practices still need to be considered and morally assessed. This makes the stages of 'data collection' and 'data acquisition' conceptually very different but ethically closely connected. Third, our selection of cases also introduces a bias: since many cases of big data use are not fully transparent, we focus on relatively well-described cases in order to have sufficient information for our systematisation. However, this information often comes as a result of investigative journalism, activism or court trials that are more likely prompted by cases that blatantly violate sensitive ethical norms. Thus, it may be that most big data use cases are significant in terms of

less inflammatory ethical concerns or are less ethically contentious in general than the cases we address.

Despite these shortcomings, the article does generate some useful insights. First, it supports the claim that ethically significant decisions are made at various stages of a big data life cycle. Although this is not a novel insight, this is the first article (in our estimation) to actually apply this logic so systematically and to do so specifically for the legal domain. Consequently, the ethics of big data practices should look beyond issues that are discretely tied to any one single stage and to scrutinise existing big data use cases along the entire data life cycle. Second, our approach offers a more structured and holistic view than what one would obtain by simply going concern by concern for a given case, potentially missing how some concerns manifest themselves in different stages and connect to one another. It thus allows us to see the prevalence of certain ethical considerations throughout the data life cycle (generalisable to the five analysed cases) and to be more thorough about how ethical concerns get compounded, alleviated or transformed throughout the life cycle.

Third, the approach adopted in this article is, in and of itself, a useful heuristic for 'lawyerly' ethical reasoning about big data use cases, which might be valuable in legal practice or in the development of big data systems. This approach can serve as a useful starting point for examining which ethical concerns tend to appear at a given data life cycle stage or even to highlight structural similarities that might be useful for developing an ethically informed typology of cases of big data practices, which could then be used to examine and address some of these cases collectively, rather than individually. From a scholarly perspective, this approach has benefits in terms of its non-committal attitude towards ethical theorising – not siding with any one particular ethical theory – offering a wider menu of ethical views for examining the morality of big data in the future. Keeping the range of ethical considerations open is arguably more conducive to fostering a discipline of big data ethics that is pluralistic and substantively richer than many of the current attempts focused on one or a limited number of master moral [.[89] This is desirable for a field of study as recent as big data ethics, where favouring any one theory or set of moral values and principles might be normatively and theoretically premature.

Finally, it bears emphasising that, despite the ethical concerns that they raise, the big data tools examined were all deemed compatible with legal norms across a variety of jurisdictions.[90] That law and morality cover distinct normative domains is hardly a surprise to most lawmakers, lawyers and legal practitioners. But here the distinction is worth recalling: since our perspective on ethics is uniquely lawyerly, the question of whether (and how) moral critiques can be recast as legal challenges remains an important area of future work for data scientists, legal scholars, legal practitioners and ethicists. The pragmatism of our approach in this article can potentially aid this recasting (given that the ethical considerations we rely on are drawn from documents that already matter for big data practices), but demonstrating the value of our approach in this respect remains a topic for further research.

89. M. Boeckhout, G.A. Zielhuis, and A.L. Bredenoord, 'The FAIR Guiding Principles for Data Stewardship: Fair Enough?', 26(7) *European Journal of Human Genetics* 931 (2018); J. Collmann & S.A. Matei, *Ethical Reasoning in Big Data: An Exploratory Analysis* (2016); D. Shin & Y.J. Park, 'Role of Fairness, Accountability, and Transparency in Algorithmic Affordance', 98 *Computers in Human Behavior* 277 (2019).

90. A special case here might be SyRI, which got through the legislative level but was finally banned at the judicial one.

Annex A

Ethical considerations[91]

Ethical consideration	Document type	Documents
Reliability	General (research conduct)	ALLEA 2017
Honesty/Integrity (qua honesty and truthfulness)	General (research conduct)	ALLEA 2017, GCC 2019, WHO 2017, UNDP 2017
Respect/mutual respect for human dignity/intrinsic value of people	General (research conduct) & Data-specific	ALLEA 2017, WHO 2017, UNDP 2017, OECD, 2016, European Commission 2019
Accountability	General (research conduct) & Data-specific	ALLEA 2017, UNDP 2017, WHO 2017, OECD 2020
Fairness	General (research conduct) & Data-specific	GCC 2019, OECD 2013, OECD 2020, European Commission 2019, EU Regulation 2016/679, CEPEJ 2018
Care (duty not to harm the subjects of research)	General (research conduct) & Data-specific	GCC 2019, ICC & ESOMAR 2007, European Commission 2019
Independence and impartiality (from bias, discrimination, prejudice and undue influence)	General (research conduct) & Data-specific	WHO 2017, UNDP 2017, UNDG 2017, OECD 2020, CEPEJ 2018
Professional commitment/Professionalism	General (research conduct)	WHO 2017, UNDP 2017
Transparency (of method and application)	General (research conduct) & Data-specific	UNDP 2017, ICC & ESOMAR 2007, UNDG 2017, OECD 2016, OECD 2013, OECD 2020, IHSN 2010, European Commission 2019, EU Regulation 2016/679, CEPEJ 2018
Explainability (as addition to transparency)	Data-specific	OECD 2016, OECD 2013, OECD 2020, IHSN 2010, European Commission 2019, EU Regulation 2016/679, CEPEJ 2018
Principled Performance/Results orientation (demonstrable benefits of the system)	General (research conduct) & Data-specific	UNDP 2017, OECD 2020, European Commission 2019
Obeying the law/Lawfulness	General (research conduct) & Data-specific	UNDP 2017, OECD 2013, OECD 2020, IHSN 2010, European Commission 2019, EU Regulation 2016/679, CEPEJ 2018
Not violating human rights	Data-specific	UNDG 2017, OECD 2020, European Commission 2019, CEPEJ 2018

91. The documents we refer to in this table are the following (ordered alphabetically): ALLEA, 'The European Code of Conduct for Research Integrity', https://allea.org/code-of-conduct/ (2017); CEPEJ, 'European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment', https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c (2018); EU Regulation 2016/679; European Commission, 'Ethics Guidelines for Trustworthy AI', https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (2019); GCC, 'Global Code of Conduct for Research in Resource-poor Setting', www.globalcodeofconduct.org/ (2019); ICC & ESOMAR, 'International Code on Market and Social Research', https://iccwbo.org/content/uploads/sites/3/2008/01/ESOMAR-INTERNATIONAL-CODE-ON-MARKET-AND-SOCIAL-RESEARCH.pdf (2007); IHSN, 'Dissemination of Microdata Files: Principles, Procedures and Practices', https://ihsn.org/sites/default/files/resources/IHSN-WP005.pdf (2010); OECD, 'The OECD Privacy Framework', www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf (2013); OECD, 'Research Ethics and New Forms of Data for Social and Economic Research', https://doi.org/10.1787/5jln7vnpxs32-en (2016); OECD, 'Recommendation of the Council on Artificial Intelligence', https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 (2019); UNDG, 'Data Privacy, Ethics and Protection: Guidance Note on Big Data for Achievement of the 2030 Agenda', https://unsdg.un.org/resources/data-privacy-ethics-and-protection-guidance-note-big-data-achievement-2030-agenda (2017); UNDP, 'Code of Ethics', www.undp.org/content/undp/en/home/accountability/ethics.html (2017); WHO, 'Code of Conduct for Responsible Research', www.who.int/about/ethics/code-of-conduct-for-responsible-research (2017).

Ethical consideration	Document type	Documents
Consent (including consent for reuse where feasible)	Data-specific	ICC & ESOMAR 2007, UNDG 2017, OECD, 2016, OECD 2013, IHSN 2010, EU Regulation 2016/679
Data minimisation (limit the collection of data to what is relevant for research)	Data-specific	ICC & ESOMAR 2007, UNDG 2017, OECD, 2016, OECD 2013, EU Regulation 2016/679
Privacy	Data-specific	UNDG 2017, OECD 2016, OECD 2013, OECD 2020, IHSN 2010, EU Regulation 2016/679
Confidentiality	Data-specific	UNDG 2017, OECD 2013, IHSN 2010
Not causing unjustified or adverse effects for individuals or groups	Data-specific	UNDG 2017, OECD 2016, European Commission 2019, CEPEJ 2018
Proportionality – Risks of harm need to be proportional to the benefits of data use	Data-specific	UNDG 2017, European Commission 2019, OECD 2016
Sensitivity to context – including focus on vulnerable population groups	Data-specific	UNDG 2017, European Commission 2019
Due diligence to evaluate data practices of third-party collaborators	Data-specific	UNDG 2017, CEPEJ 2018
Data and analysis quality assessments (general and to prevent biases)	Data-specific	UNDG 2017, OECD 2016, European Commission 2019, EU Regulation 2016/679, CEPEJ 2018
Sharing data (to the extent it does not violate other principles)	Data-specific	OECD 2016
Responsibility to maintain adequate security of data	Data-specific	OECD 2013, OECD 2020, EU Regulation 2016/679, CEPEJ 2018
Promotion/adherence to democratic values and individual freedom	Data-specific	European Commission 2019, OECD 2020
Not limiting user autonomy	Data-specific	CEPEJ 2018